

CHANCE AND THE ORIGIN OF LIFE

EDWARD ARGYLE

Dominion Radio Astrophysical Observatory, National Research Council Canada, Herzberg Institute of Astrophysics, Penticton, B.C., Canada V2A 6K3

(Received 12 July, 1977)

Abstract. Random chemical reactions in the Earth's primitive hydrosphere could have generated no more than 200 bits of information, whereas the first Darwinian organism must have encoded about a million bits, and therefore could not have arisen by chance. This information gap is bridged by separating reproduction from organism, and postulating a reproductive chemical community that would generate information by proto-Darwinian evolution. The information content of the initial community could have been as low as 160 bits, and its evolution might have led to the first Darwinian cell.

1. Introduction

It is a widely held view that life will arise spontaneously on the surface of any planet that provides a suitable physical and chemical environment. This belief is saved from tautology by the generously broad definitions of 'suitable' that abound in discussions of the origin of life. Indeed it is almost sufficient to require only that liquid water occur on the planet's surface, for then it follows that the atmospheric pressure and ambient temperature will be in ranges that promote a rich variety of organic reactions.

On the ancient Earth, as today, the simultaneous presence of the three states of matter along terrestrial shorelines provided reaction sites and macroscopic transport for most of the planet's chemicals. The temperature was low enough to confer a substantial lifetime on thermodynamically improbable molecules formed in sunlit waters, yet high enough to give speed to the processes of chemistry, and to the evolution of life. The importance of speed in chemistry and evolution is emphasized by the reflection that a cooler planet than ours, where reaction rates were one fourth as great, would see its sun burn away from the main sequence of stars before it witnessed intelligent life.

This standard scenario of life's origin has been strengthened greatly by the outcome of laboratory experiments in geochemistry (Fox and Dose, 1972). They show that the assumed primitive molecules of our planet's early atmosphere, if supplied with free energy, could form sugars, amino acids, purines, pyrimidines and other life-related organic substances. More recently, the discovery of numerous interstellar organic molecules by radio astronomers (Robinson, 1976) has all but confirmed that organic molecules of some complexity are of widespread occurrence in the universe. Hence it is now widely believed that the primordial Earth was the scene of a varied organic chemistry that was rescued from the death of equilibrium by the energy flow from the sun.

At this point in the development of life on Earth the scene goes briefly out of focus.

'Somehow', it is thought, the first very simple, but living reproducing cell was formed. The picture then clears as Darwinian evolution steps in and drives the progeny of that first cell through inexorable stages of experiment and success to a culmination in intelligent life.

It is the purpose of this paper to discuss the events that took place between the settling of the Earth's crust about four thousand million years ago and the appearance of the first living organisms a few hundred million years later. The approach is through information theory, for the light that it throws on the probabilities of random formation of molecules that might have been relevant to the origin of life on Earth.

2. Information in Sequences

Information theory has been developed to assist in the design of communication channels and the codes used with them. As such the theory deals with the mathematical properties of a message, or sequence of symbols, and is not concerned with *meaning*. However, two concessions are made with respect to meaning. If the message is without significance to the receiver, or if it has been received before, it is conceded to contain no information.

The definition of the information content of a message is constructed to meet two intuitive needs. The amount of information should be proportional to message length, and it should increase appropriately with the richness of the alphabet in which it is encoded. The more symbols there are in the alphabet the greater is the number of different messages of length n that can be written in it. The number of permutations of k alphabetic symbols, taken n at a time is

$$N = k^n, \quad (1)$$

where N is the number of different possible messages. The information content, H , of a significant message is defined to be

$$H = \log N. \quad (2)$$

If the logarithm is taken to the base 2 (as always in this paper) the result is expressed in binary units, or bits.

The probability that a random sequence of n symbols would convey the intended message exactly is only $1/N$. In this sense a significant message is an improbable event. If improbability is taken to be the reciprocal of probability, formula (2) states that the amount of information in a message is simply the logarithm of its improbability of occurrence.

An alternative formula for H can be obtained by substituting (1) into (2) yielding

$$\begin{aligned} H &= \log k^n, \\ &= n \log k. \end{aligned} \quad (3)$$

Clearly, the information content of a message is proportional to its length, and also to the logarithm of the number of symbols in the alphabet it uses. In fact, $\log k$ is the

information content per symbol. Because (3) is linear in n , the H -values of several messages received in succession (like one longer message) are additive.

The strict validity of these equations requires the condition that each of the k symbols be equally likely to occur anywhere in the sequence. In other words, the code must be non-redundant. This assumption is not seriously violated in the circumstances that will be examined.

That H is a conveniently compact measure of information is seen when formulas (1) and (2) are applied to a biochemical sequence such as messenger ribonucleic acid (mRNA). Assuming that an average gene comprises a string of $n = 1200$ nuclear bases of $k = 4$ kinds, the number of different genes of that length is, by (1), the enormous number

$$N = 4^{1200} = 2^{2400} = 10^{722.5};$$

whereas the information content, using (2), is

$$H = \log 2^{2400} = 2400 \text{ bits,}$$

a manageably small number.

Apart from purposes of illustration it is easier to arrive at this result by using a form of Equation (3). Since $k = 4$, $\log k = 2$ and therefore

$$\begin{aligned} H_{\text{RNA}} &= 2n, \\ &= 2400 \text{ bits.} \end{aligned} \tag{3a}$$

There are $k = 20$ different amino acids in a modern peptide chain, or protein. Therefore $\log k = 4.32$ and (3) becomes

$$H_{\text{PEP}} = 4.32n \tag{3b}$$

for a chain of n amino acids.

An important property of communication channels can now be brought out by comparing the information content of a typical protein with that of the gene that specifies it. Because of the triplet nature of the genetic code the 1200-base gene considered above has 400 codons and therefore is translated into a protein of 400 amino acids. But, by (3b) the information content of the protein is only $4.32 \times 400 = 1729$ bits. About 671 bits have been lost in translation. This feature of genetics is known as the *degeneracy* of the genetic code (64 codons represent only 20 amino acids plus start and stop signals). More physically, it is a degeneracy of the translation process. Clearly, a cell equipped with about 60 different transferase molecules (tRNA) instead of 20, could utilize 60 different amino acids for protein manufacture — all without any change in the basic structure of the mRNA. Degeneracy may have been important for the origin of life, and will be mentioned again.

A useful 'benchmark' to the information content of living organisms is provided by the common bacterium *Escherichia coli* (*E. coli*). Assuming it to have 2500 genes it will encode a total of six million bits. The amounts of information stored in this a 1 ther relevant structures are displayed in Table I.

TABLE I
The information content of various structures

Structure or sequence	Number of different possible cases or states	Amount of information, bits
A molecule that is certainly present	1	0.
Toggle switch	2	1.
Nuclear base	4	2.
Amino acid	20	4.32
Combination lock	10^6	20.
Modern tRNA molecule	$10^{4.5}$	150.
Random experiments	$10^{6.3}$	200.
Gene or Protein	$10^{7.22.5}$	2400.
Virus (50 genes)	$10^{36\ 000}$	1.2×10^5
E. coli (2500 genes)	$10^{1\ 800\ 000}$	6.0×10^6
Man (100 000 genes)	$10^{72\ 000\ 000}$	2.4×10^8

At this point it is important to make clear the nature of the argument from information theory as it applies to the origin of the first living organism. We are willing, perhaps, to make the optimistic assumption that the formation of random peptide chains of amino acids was thermodynamically favorable, or at least permitted, thanks to the presence of the necessary organic substances, mineral catalysts and available energy. In other words, random peptides were commonplace. But a random chain of n amino acids contains no information unless it has significance for the origin of life. If it, alone amongst such chains, has that significance, its information content will be given by formula (2) with N set to the number of different possible chains of length n . (Equally, depending on one's predilection regarding the manner in which life started, the formation of a significant nucleic acid would generate information.)

If life could have started in but one way it would be possible, in principle, to deduce how it began, given only that there is life. Such an origin would therefore have been, in the language of physics, 'nondegenerate'. But if two or more distinct sequences of amino acids (or bases) could have independently triggered the process, then the origin of life would have been a degenerate event. The mere knowledge that life exists would no longer suffice to deduce the actual mode of origin out of the several that were possible. Clearly, life on earth today is a highly degenerate process in that there are millions of different gene strings (species) that spell the one word, 'life'.

3. The Generation of Information

If the first cell was as complex as *E. coli*, a momentary suggestion made only for illustration, it would have been necessary for our planet's early random chemistry to have generated about six million bits of information by pure chance. The way in which chance experiments can generate information is now examined by looking at a game.

Consider a locked door equipped with a combination lock comprising twenty

toggle switches on its outer surface. If all twenty switches are set correctly the door can be opened by grasping and turning the handle. A person having no knowledge of the correct combination can open the door by making a (long) series of random experiments. A few switch positions are changed at random and the handle is tried. This procedure is repeated until the door opens. The experimenter can now write down the combination by looking at the successful switch positions. Clearly, he has acquired 20 bits of information, 1 bit per switch, by making, probably, about one million random experiments. (There are $N = 2^{20} \approx$ one million different possible switch combinations.)

Here again it is seen that a random sequence (of switch positions) is uninformative as long as it is undistinguished. But as soon as one sequence proves to be significantly different from the other million, it carries 20 bits of information. The average amount of information H_r , that can be generated by N random experiments is seen to be

$$H_r = \log N, \quad (4)$$

and this is merely a specialization of Equation (2).

To say that random experiments can *generate* information is perhaps a subjective view not warranted by the facts. It could be argued that the information necessary to open the door is already encoded in the pattern of the electric wiring that connects the switches to the lock mechanism, and that all one need do to acquire that information is to look at the other side of the door! Similarly, it will be true that the sequence of units necessary to the first reproductive structure was already encoded in some deep way in the environment that gave it birth. This may be what the philosophers mean when they say that the potential for life inheres in the very nature of matter.

4. The Probability of the Origin

In the standard scenario the Earth's fluids were in a continual turmoil of chemical change that produced organic sequences in a random way. All but one of these long molecules were passive. The unique molecule quickly organized most of the carbon in its environment into copies of itself. The information necessary to reproduce had been transferred from the chemical milieu in which it had lain concealed to a molecule capable of recording it. The random chemistry was then replaced by Darwinian evolution.

Calculation of the probability that a reproductive chemical was formed by chance requires a knowledge of the rate at which relevant random reactions occurred, and the length of time they continued. Neither figure is well established — especially the first — and no realistic calculation of the information-generating power of the environment can be made. On the other hand it is possible to set optimistic upper limits to the reaction rate and the available time, and then to derive the maximum possible value of H_r permitted under the assumptions made.

Assuming that the early waters of Earth contained 10^{44} carbon atoms (Suess, 1975), it is optimistic to take the number of amino acid molecules at 10^{43} . If these molecules were linked in random sequences of average length 10 there would be about 10^{42} peptide

chains. If the mean lifetime against extension or breakage for the average chain were 10 milliseconds, and the action continued for 500 million years, the total number of peptides formed would be

$$(10^{42} \text{ peptides}) \times (5 \times 10^8 \text{ yrs}) \times (3 \times 10^7 \text{ s yr}^{-1}) \\ \times (100 \text{ reactions s}^{-1} \text{ peptide}^{-1}) = 1.5 \times 10^{60}.$$

However, not all of these are different peptides. For the case in which there are 20 kinds of amino acids it can be shown that 98.2% of the 1.5×10^{60} reactions are repeat productions of short peptides. Subtracting these, there remain

$$N = 2.75 \times 10^{58}$$

different molecules, and the corresponding information content is, by (4)

$$H_r = 194 \text{ bits.}$$

If the calculations are made for nucleic acids instead of peptides the result is substantially the same.

It would seem impossible for the prebiotic Earth to have generated more than about 200 bits of information, an amount that falls short of the 6 million bits in *E. coli* by a factor of 30 000. A natural attempt to save the scenario is to postulate a simpler first cell. However, there is little to be gained through this proposal. An average virus codes about 2% as much information as *E. coli* (120 000 bits) and is not capable of reproducing in an abiotic environment. Rather it must subvert the metabolic machinery of a regular cell for materials, energy and protein synthesis. It is difficult to imagine an independently reproductive cell as simple as a virus (Watson, 1970–1), and even if one can, it helps little to bridge the enormous information gap between chemistry and life.

Parenthetically, it is interesting to note that if the probability of the chance appearance of life on Earth seems remote, there is little comfort to be gained by enlarging the arena to the whole galaxy. Even if there are 10^9 Earth-like planets in the Milky Way the potential for random generation of information rises only to 224 bits – less than 0.2% of the content of the average virus.

Even one gene of average length encodes about 2400 bits, so it is not useful to speak of a primitive naked gene that reproduced unless it was so short that it specified a protein of no more than about 33 amino acids. Whether one prefers to think of the first nucleic acid, the first gene, the first protein or the first enzyme as the unique structure that began life, there is the difficulty of visualizing the way so small a molecule could have commanded the environment to its selective reproduction.

If life on Earth had a spontaneous origin there must have been an intermediate mechanism that was capable of augmenting the information content of one or a few early molecules up to the million-bit level required by the first organism. But before turning to a possible new mechanism it is worthwhile to consider how Darwinian evolution is able to generate information so much faster than random experimentation. The question is philosophically interesting because, as stressed in every good introduction to the subject,

the mutations that provide the raw material of evolution occur at random and are made without purpose or goal. Also, the random experiments, mutations, occur very slowly, about one per generation, and require a great deal of carbon per individual – about 10^{10} atoms. If most of the ocean's carbon resided in cells like *E. coli*, reproducing once per hour for three billion years, the total number of mutations would be only about 10^{47} , less than 10^{-13} of the number allowed for random chemical experiments. Nevertheless Darwinian evolution has been able to amass prodigious amounts of information in the world's living species. How this was done can be seen by examining a modified version of the game with the locked door.

5. Darwinian Generation of Information

Consider now a slightly different kind of door – one that opens a millimeter for each switch that is set correctly. The initial switch positions are random. By noticing the door's response, the uninformed operator can open it very quickly. He selects a switch at random and alters its setting. If the door opens a little he makes another experiment. Otherwise he returns the switch to its original (correct) position and then goes on to the next experiment. Twenty systematic trials would have opened the door, but about 60 are necessary if they are random because of needless experimenting with switches already tried.

This door-opening game can be seen to correspond to an extremely simplified model of Darwinian evolution if certain identifications are made. The door is the reproducing organism and its openness is a measure of its success. In each experiment (generation) the door is seen in two states – its original state, and its new state after a switch is thrown. These are the two progeny of a divided cell, one normal and the other mutant. Only the 'fitter' of the two states is tolerated by the operator (the environment), which then destroys the other state (offspring) and leaves the former to serve as original state (parent) for the next experiment (generation).

The 20 bits of information required to open the Darwinian door have been generated in only about 60 experiments (generations), instead of the million required to open the random door treated earlier. Thus

$$H_D = cN, \tag{5}$$

where H_D is the amount of information generated by N Darwinian experiments, and c is a factor that equals $\frac{1}{3}$ in this example.

Of course, it is not accurate to assume that every mutation that makes the base sequence of a gene more like that of a superior gene will necessarily itself be an improved gene. In other words the shortest route to a better gene will not necessarily thread a series of monotonically improving genes. Therefore the formula for H_D gives an upper limit rather than a precise estimate of the amount of information generated in N experiments. Furthermore, the factor $\frac{1}{3}$ is appropriate only to the game described. It will be smaller for longer genes carrying more than 20 bits, for mutation rates less than the 50% used in the

game, for larger populations (because of duplicate mutations) and for nearly perfected genes that are harder to improve. Unpublished computer simulations suggest a c -value in the range 10^{-2} to 10^{-6} for plausible simple model organisms.

Even the lower values of c have little impact on the disparity between H_D and H_r for interesting values of N . When $N = 10^{40}$, $H_D = 10^{34}$ and $H_r = 132$. The Darwinian organism acts like a machine for generating information. Its special function is to copy all its genes, including those carrying random alterations to the message. All new messages are then subjected to environmental scrutiny. It is immaterial that only rarely does a new message pass the test. Once approved, it is copied at an exponentially increasing rate and, for a time, becomes the 'standard' message that underlies future attempts to encode even more information about the environment's tolerance for life. This process works because the reproductive power of a population of organisms exceeds the environmental culling that takes place between generations. The cost imposed by genetic experimentation is paid out of surplus reproduction. Meanwhile mutations that are not rejected add information to the genes at the Darwinian rate.

6. An Intermediate Mechanism

Any proposed new mechanism for generating information faster than possible by pure chance faces a fundamental *logical* difficulty if it is embodied in an integrated structure. Any structure that reproduces at a rate that outruns decay processes will undergo Darwinian evolution, and for that very reason will be a Darwinian organism. This dilemma can be avoided by dropping the tacit assumption that reproduction is possible only to an integrated structure such as a cell or other living organism. If an amorphous *community* of free molecules could reproduce it could also evolve in a proto-Darwinian way, but would not be living because it would contain no structure that independently reproduced itself.

Although a cell such as *E. coli* would not reproduce in today's environment if it lost its wall there seems to be no impediment to the general idea that the contents of a broken cell (along with the wall fragments), when poured into a benign environment might still carry on the biochemical reactions characteristic of reproduction (Horowitz, quoted in Margulis, 1970). If that happened there would be a multiplication of the populations of biochemicals at the expense of the simpler organic precursors in the region. Such a system would contain no living entity but certain molecules in it would make copies of each other even though no molecule reproduced itself. It must now be asked whether one can postulate a chemically reproductive community containing less than 200 bits of information without forsaking plausibility.

7. A Reproductive Chemical Community

An important requirement for plausibility in any proposal for the origin of life is the recalculation of H_r on a less wildly optimistic basis. To this end the assumed concentra-

tion of amino acids (or nucleotides) is reduced by the factor 1000 to 10^{40} in all the oceans; the reacting mass is reduced from the entire hydrosphere to those waters shallow enough to permit sunlight to reach the bottom, a factor of 100; and the lifetime of an average molecule is increased to 1000 seconds, a factor of 10^5 . The product of these factors is 10^{10} and the lost information entailed by their use is 33 bits. Thus a more plausible value for the amount of randomly generated information is

$$H_r = 161 \text{ bits.}$$

The idea of a reproductive chemical community can be made concrete by adopting the suggestion of Crick *et al.* (1976) that early protein synthesis took place without the use of even a simplified ribosome. Their scheme achieves translation of a primitive genetic code to protein by means of transfer RNA molecules (tRNA) each carrying a 7-base anticodon loop at its 5' end and a specific amino acid-recognition group at the 3' end. An attempt will now be made to calculate the minimum amount of information required to create the necessary molecules.

In its simplest and most cogent form the new mechanism recognizes the four amino acids glycine, asparagine, serine and aspartic acid, whose modern codons occur in the bottom right hand corner of the codon table. Consequently, four tRNA's are involved in the translation of a mRNA molecule. Because of the degeneracy of the 3-base codon only one bit is required for each base in the mRNA, and therefore also for each of the three variable bases in the 7-base anticodon. The other four fixed anticodon bases are nondegenerate and represent two bits each. Thus 11 bits are required for each of the four 7-base anticodons.

The problem of amino acid recognition is not dealt with by Crick *et al.* Presumably each tRNA forms a recognition cavity at its 3' end. Allowing 8 bases for the cavity would call for 16 bits for each tRNA. However, the occurrence of considerable degeneracy, especially in primitive structures, is widespread in the biochemical world (Hasegawa, 1975) and it may not be unreasonable to reckon 12 bits per cavity. If functional tRNA's could be this simple they would embody only 23 bits each.

Only one molecule remains — the mRNA, and it will be assigned the remaining information. Subtracting 4×23 from the allowed 161 bits leaves 69 bits for a mRNA molecule containing 23 3-bit codons. Without further appeal to degeneracy it is now entirely proper to demand that the mRNA specify the best nonspecific RNA replicating enzyme possible out of all peptides composed of 23 amino acids, each chosen from the set of four.

At this stage in the development of the idea of a reproductive chemical community it would be all too easy to conclude that the replicase would form copies of all 5 RNA's while they were busy making more replicases. Indeed, the formation of the first replicase *starts* promptly because the principle of additivity of information used in this treatment applies only to the case where all the specified molecules are made at the same time, in the same place and in specified relationship to each other. That is, the 161 bits of information include specification of the correct initial juxtaposition of the 5 molecules

to *begin* translation. But translation will not be completed unless the used tRNA's that drift away from the mRNA after the formation of each peptide bond are brought back *again and again* by Brownian diffusion until all 22 bonds have been made. Furthermore, all this must happen during the postulated 1000-second lifetime of the molecules. To ensure completion of the first replicase it may be necessary to suppose that the crucial molecules were trapped in a micrometer-sized interstice between particles of clay or other material. After a substantial population of each of the six molecules had been generated the system would become secure against diffusion, and would no longer require, or benefit from, confinement in a small volume.

8. Proto-Darwinian Evolution

In the modern cell mRNA codons can be translated into amino acids at the rate of 40 per second, with an error ratio below one in a thousand. Replication of nucleic acids is about equally fast, and nearly a million times more accurate (Watson, 1970–2). In the first reproductive chemical community envisaged here the population would double in every generation if about two reasonably accurate sets of six molecules were produced every 1000 seconds, per mRNA. Thus a high mutation rate combined with sluggish replication and translation would be permitted, as well as expected. If such a system generated information at the rate of 0.1 bit per mutation, ($c = 0.1$ in formula (5)), only a few hours would be required to double the original investment of 161 bits. Thus the onset of reproduction and proto-Darwinian evolution in a randomly generated system that might have taken hundreds of millions of years to arise marks an exquisitely critical point in its history.

In a proto-Darwinian community there are no organisms to compete with each other. The information necessary to reproduction is dispersed in several free RNA molecules. (The central dogma of Darwinian genetics, that information flows from nucleic acid to protein, but not along the reverse path, is already visible in the fact that the peptide molecule, replicase, is not an essential element of the reproductive set.) Tolerable errors in translation or replication lead not to new organisms but to new chemical pathways, and these compete. Because of the slowness of long range molecular diffusion, competition is mostly a local matter. A novel pathway can survive for a time in its own region even though it is not the best in the pool. But in the long run the pool evolves as a whole.

Each mRNA molecule acts as a centre of translation, accepting all appropriate tRNA's that drift into contact with it. Thus each replicase is assembled by random tRNA's in the community surrounding the mRNA. It is noteworthy that the system of Crick *et al.*, as developed here, posits more information in the tRNA's than in the mRNA. If the five RNA's are all regarded as genes it could be said that proto-Darwinian reproduction is hyper-sexual in the sense that nucleic acids mix even more freely than in sexually pairing organisms.

Assuming that the reproductive chemicals eventually leak into new niches and evolve complexity and diversity, there would soon be a variety of specific pathways plied by

separate communities of chemicals of ever-increasing molecular weight and specificity. However proto-Darwinian systems lack one important feature of Darwinian populations – safeguards against hybridization. Two divergently evolved chemical communities spilling into a common pool would unavoidably hybridize if their pathways contained any common segments, as would seem very likely. For example, if there had been divergent evolution of genetic codes the indiscriminate mixing of nucleic acids would thwart the interlocking specificities so laboriously built up over their histories and would waste resources prodigiously. This difficulty would dog the path of proto-Darwinian evolution until some means of collecting and sheltering a reproductive set of chemicals was evolved. Perhaps the Darwinian organism was evolution's answer to the problem of hybridization.

It is easy to suggest ways in which a reproductive set of chemicals might clump together in relative isolation from the rest of the community. The difficulty is to have the isolation partial in just the way that permits essential precursor molecules to enter the enclave but prohibits the loss of genetic material until it can be discharged in a self-reproductive clump. *That this partial isolation is not easy to specify is just the problem of the origin of life.* It probably required a million bits of information, and it is the burden of this paper that that information might have been generated by the proto-Darwinian evolution of a reproductive community that began with less than 200 bits of randomly generated information.

The shortcomings of this scheme to start some kind of rapid information-generating process in the prebiotic soup are too obvious to ignore. The broth is speciously thick and the prospects of the first community are precarious in the extreme. It would be helpful to know more about the catalytic powers of short peptides, especially those containing only two kinds of amino acids. Such enzymes could be assembled by two tRNA's. If the RNA replicases turn out to be highly degenerate, less information will be needed for their formation. These and other possibilities remain to reduce the improbability of the chance occurrence of a reproducing system.

9. Conclusions

Improbable structures can be formed by random trials if the latter are sufficiently numerous. Information theory simplifies the task of separating the possible from the impossible by reducing structural complexity and experimental prodigality alike to a common informational measure expressed in bits.

The calculation of the information-generating power of the Earth's primitive hydrosphere offered here is neither precise nor definitive. Rather it is suggestive that there is an enormous information gap between the products of a random chemistry and the simplest imaginable reproducing organism.

It seems futile to force Darwinian evolution backwards through simpler and simpler organisms to one whose structure could have been the outcome of random trials. Instead it is proposed that special molecules that arose by chance formed a reproductive

community of sufficient vigor to start a proto-Darwinian evolution that dominated its development.

Proto-Darwinian evolution will have been significant for the origin of life if at least one reproductive chemical community can be specified by not more than 200 bits of information, and does not lie in an evolutionary cul-de-sac.

If the 200-bit figure is seriously in error, it is too large. If the true figure is less than half this upper limit it will probably be necessary to discover information-generating mechanisms beyond those discussed here. Alternatively, it would be encouraging to discover that enzymes are highly degenerate molecules that economize on information.

Note added in proof: The prospects of the reproductive chemical community described in Section 7 become more promising when it is noted that the system is slightly degenerate. If the specifications of the five molecules are thought of as five words in a 161-bit message it is seen that there will be $5!$ permutations of word order, and that these do not change the meaning. Moreover, there are $4!$ possible sets of associations between the anticodons and the amino acid recognition sites of the four tRNA's. Although each such set would require a different mRNA to specify the replicase molecule, the appropriate messenger can always be encoded by the 69 bits allotted to it. Consequently, the postulated system would arise $4! \times 5! = 2880$ times at diverse places in the hydrosphere. Only one of these systems need have propagative success to start proto-Darwinian evolution.

References

- Crick, F. H. C., Brenner, S., Klug, A., and Piecznik, G.: 1976, *Origins of Life* 7, 389.
- Fox, S. W. and Dose, K.: 1972, *Molecular Evolution and the Origin of Life*, W. H. Freeman and Co., Chap. 4.
- Hasegawa, M. and Yano, T.: *Origins of Life* 6, 219.
- Margulis, L.: 1970, *Origins of Life*, Gordon and Breach, New York. p. 311.
- Robinson, B. J.: 1976, *Proc. Astron. Soc. Aust.* 3, 12.
- Suess, H. E.: 1975, *Origins of Life* 6, 9.
- Watson, J. D.: 1970-1, *Molecular Biology of the Gene*, W. A. Benjamin Inc., New York, p. 503.
- Watson, J. D.: 1970-2, *Molecular Biology of the Gene*, W. A. Benjamin Inc., New York, pp. 297 and 368.